Genomics 95 (2010) 73–83

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

EBV transformation and cell culturing destabilizes DNA methylation in human lymphoblastoid cell lines

D. Grafodatskaya [a], S. Choufani [a], J.C. Ferreira [a], D.T. Butcher [a], Y. Lou [a], C. Zhao [a], S.W. Scherer [a,b], R. Weksberg [a,c,*]

[a] Program in Genetic and Genomic Biology, Hospital for Sick Children Research Institute, 101 College street, Toronto, Ontario, Canada M5G 1L7
[b] The Centre for Applied Genomics, The Hospital for Sick Children, 101 College street, Toronto, Ontario, Canada M5G 1L7
[c] Division of Clinical and Metabolic Genetics, Hospital for Sick Children, 525 University Avenue, Toronto, Ontario, Canada M5G 2L3

ARTICLE INFO

Article history:
Received 19 June 2009
Accepted 1 December 2009
Available online 18 December 2009

Keywords:
DNA methylation
Lymphoblastoid cell lines
White blood cells

ABSTRACT

Recent research suggests that epigenetic alterations involving DNA methylation can be causative for neurodevelopmental, growth and metabolic disorders. Although lymphoblastoid cell lines have been an invaluable resource for the study of both genetic and epigenetic disorders, the impact of EBV transformation, cell culturing and freezing on epigenetic patterns is unknown. We compared genome-wide DNA methylation patterns of four white blood cell samples, four low-passage lymphoblastoid cell lines pre and post freezing and four high-passage lymphobastoid cell lines, using two microarray platforms: Illumina HumanMethylation27 platform containing 27,578 CpG sites and Agilent Human CpG island Array containing 27,800 CpG islands. Comparison of genome-wide methylation profiles between white blood cells and lymphoblastoid cell lines demonstrated methylation alterations in lymphoblastoid cell lines occurring at random genomic locations. These changes were more profound in high-passage cells. Freezing at low-passages did not have a significant effect on DNA methylation. Methylation changes were observed in several imprinted differentially methylated regions, including DIRAS3, NNAT, H19, MEG3, NDN and MKRN3, but not in known imprinting centers. Our results suggest that lymphoblastoid cell lines should be used with caution for the identification of disease-associated DNA methylation changes or for discovery of new imprinted genes, as the methylation patterns seen in these cell lines may not always be representative of DNA methylation present in the original B-lymphocytes of the patient.

Introduction

Immortalized lymphoblastoid cell lines (LCLs) are of great practical value for human genetics, because they provide a virtually unlimited source of DNA. Many publicly available research repositories store LCLs derived from both healthy individuals and individuals with various disorders. These cell lines have been particularly useful for genetic studies despite rare cases of genomic instability [1].They have also been successfully used for gene expression analyses in neuropsychiatric disorders [2,3]. Given the attention focused now on discovery of epigenetic determinants in human disorders, we studied the stability of epigenetic marks in LCLs, easily accessible cell lines for such research.

DNA methylation is an important epigenetic mechanism regulating gene expression. It occurs predominantly at cytosines of CpG dinucleotides. CpGs are distributed non-randomly in the genome. For 98% of the genome, CpG sites occur at low frequency once per 80 dinucleotides. CpG islands range in size from 200 bp to several kb, comprise 1–2% of the genome, and have a five-fold higher concentration of CpGs than the rest of the genome [4]. The CpG islands commonly associated with gene promoters are usually not methylated. Important exceptions are the imprinted genes and genes on the inactive X chromosome [4]. Normally methylation of promoters containing CpG islands causes gene silencing. However, for CpG poor promoters there are conflicting reports about the effect of methylation on gene expression [5,6].

Aberrant DNA methylation can result in various disorders including cancer [7] and imprinting disorders [8,9]. Whole genome based approaches such as microarrays and deep genome sequencing have presented new opportunities to identify epigenetic defects associated with human diseases. In many cases, DNA from LCLs might be considered the best available material for such experiments. However, the essential question is: normal methylation marks retained in LCLs? There are several factors that could potentially change methylation patterns in LCLs as compared to B-lymphocytes, from which they originate. These include Epstein Barr virus infection and subsequent immortalization, cell culture conditions and freezing cycles. Two types of altered methylation patterns could theoretically occur. *De novo* methylation of normally unmethylated sequences could occur via DNA methyltransferase activity. Alternatively, loss of methylation could occur either passively through failure of DNA

* Corresponding author. Division of Clinical and Metabolic Genetics, Hospital for Sick Children, 525 University Avenue, Toronto, Ontario, Canada M5G 2L3. Fax: +1 4168135345.
E-mail address: rosanna.weksberg@sickkids.ca (R. Weksberg).

0888-7543/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2009.12.001

methylation maintenance following DNA replication cycle, or actively by demethylating activity; however no enzyme with such activity has been identified to date. If DNA methylation changes were to occur in LCLs at specific genomic locations, this would not be a problem for epigenetic research, as the probes for these genomic regions could be excluded from the analysis. However, if methylation changes occur at random genomic locations in LCLs, this could invalidate experiments targeting the identification of candidate DNA methylation changes associated with a specific disease. This is because it would not be clear whether the altered DNA methylation pattern represents the original patient's DNA profile or whether it reflects in vitro changes that occur during cell line establishment and maintenance.

Our experiments were designed to explore the scope of epigenetic alterations occurring in LCLs and how that might impact the use of these lines for epigenetic discoveries. We compared genome-wide DNA methylation profiles from LCLs and white blood cells (WBCs) using two technologies: methylated DNA immunoprecipitation followed by Agilent CpG island array (27,800 CpG islands) and Illumina's Infinium HumanMethylation27 BeadChip (27,578 individual CpG dinucleotides spanning 14,495 genes). Our goal was to determine the utility of these two platforms to identify methylation changes in comparisons of methylation profiles of WBCs and LCLs. We compared four WBC samples, four low passage LCLs (LP_LCLs) and four low passage LCLs after freezing (FR_LCLs) from the same individuals and four unrelated high passage previously frozen LCLs (HP_LCLs). We used both Illumina and Agilent platforms for genome-wide correlation analysis and analysis of DNA methylation changes in imprinted genes. The Agilent platform was used for differential methylation analysis of one WBC and LP_LCL pair in order to select regions of several hundred bps with methylation changes, as except for imprinted genes Illumina coverage was limited to two CpG sites for most genes. The Illumina array and pyrosequencing were used to validate differences detected by Agilent in a larger number of LP_LCL pairs.

Our results show that the various technologies we used validate each other in identifying DNA methylation changes. We found random methylation changes in LP_LCLs, FR_LCLs and HP_LCL; the largest frequency was in HP_LCLs. The methylation profiles in FR_LCLs were similar to LP_LCLs, indicating that one cycle of freezing does not induce significant methylation changes.

We also checked DNA methylation patterns at differentially methylated regions of imprinted genes. Imprinted genes are expressed from only one of the parental alleles. There are approximately 60 imprinted genes found in the human genome (http://www.geneimprint.com). There are a number of single imprinted genes (e.g. DIRAS3, NAP1L5, NNAT) but the majority identified to date are clustered in the genome [10]. The parent-of-origin specific expression of imprinted genes within the cluster is usually regulated by allelic methylation at a single imprinting centre (IC), and often additional differentially methylated regions (DMRs) are present within the cluster, usually within the promoters of the imprinted genes. These parent of origin specific allelic methylation patterns are maintained in somatic cell lineages, with few exceptions [8]. However, it is not known whether this differential allelic methylation is consistently maintained in immortalized LCLs. Our results showed that several imprinted DMRs, but none of the known ICs, have methylation changes. These occur in all type of LCLs, with more changes occurring in HP_LCLs.

Results

We compared DNA methylation between WBCs and LCLs at two levels of resolution: CpG islands and single CpG sites, using two microarray platforms, the Illumina HumanMethylation27 BeadChip and the Agilent CpG island array. For the Agilent array, the methylated DNA immunoprecipitation (MeDIP) was carried out prior to hybridization.

MeDIP is a recently developed technique [11] used to capture methylated DNA sequences. We labeled the immunoprecipitated DNA and input DNA with two different fluorescent dyes and co-hybridized to the Agilent microarrays. Log2 ratios of immunoprecipitated DNA versus input (MeDIP/Input) were used to assess the levels of methylation with higher log2 ratios representing higher levels of methylation. The log2 (MeDIP/Input) is not an absolute methylation value, but can be used to compare methylation levels between samples at the same genomic locations (i.e. microarray probes), but not to compare methylation levels among different genomic locations, as MeDIP enrichment depends on the density of the methylated CpG sites [11,12]. The Illumina platform utilizes bisulphite converted DNA and thus measures absolute methylation levels at specific CpG sites. The methylation level of each site is measured as $C/(C+T)$ and is represented as a β value ranging continuously from 0 (0% methylated) to 1 (100% methylated).This method provides more accurate methylation level measurements than MeDIP, but less genomic coverage as only one CpG site is measured by each array probe.

We implemented both of the above described techniques in order to study DNA methylation stability in LCLs. We assessed changes in DNA methylation for DNA fragments 500 bp and longer by MeDIP and Agilent CpG island microarray and the DNA methylation changes occurring at specific CpG sites by Illumina methylation platform.

To assess the methylation differences between DNA from WBCs and LCLs, we analyzed methylation profiles for four sets of WBCs, low passage LCLs, and frozen LCLs from two males and two females (WM1/LM1_LP/LM1_FR, WM2/LM2_LP/LM2_FR, WF1/LF1_LP/LF1_FR, WF2/LF2_LP/LF2_FR) using the Illumina platform, and one pair of WBC and low passage LCL (WM1/LM1_LP) using MeDIP and the Agilent platform (Table 1). DNA was extracted from the low passage LCLs directly after EBV transformation prior to freezing the cells, and then DNA was extracted a second time after one cycle of freezing (frozen LCLs). Four high passage LCLs (LM3_HP, LF3_HP, LM4_HP, LPID11_HP) had undergone an unknown number of freezing cycles and cell divisions, and were studied using the Illumina array; and two of them (LF3_HP and LPID11_HP) using the Agilent array. As well, one un-paired WBC sample (WF4) was studied with the Agilent array (Table 1).

Comparison of genome-wide DNA methylation levels between LCLs and WBCs

DNA methylation across the genome was assessed to determine consistency among WBCs versus LCLs. We first assessed cell type specific differences between WBC and three types of LCLs by performing Manhattan hierarchical clustering analyses of Illumina methylation data. Clustering showed that WBC samples and LCLs cluster in separate groups, whereas there was no clustering of LCLs according to the type (low passage, frozen, high passage) (Fig. 1). This suggests that DNA methylation changes in each group could be either specific for the group or random.

To further assess cell type specific differences, we compared genome-wide DNA methylation profiles between individual samples by calculating correlation coefficients (R) for linear regressions of pair-wise sample comparisons. Sex chromosomes (X and Y) were excluded from these analyses, in order to pick up only cell type specific differences, but not sex specific differences.

Illumina methylation correlations

Correlation coefficients (R) for Illumina methylation are shown in Table 1A. R values among WBC samples ranged from 0.985 to 0.99 (Table 2A), showing very small inter-individual variation in the DNA methylation of WBCs. However, for pairs of WBCs and LCLs from the same individuals, R values were in range of 0.933–0.947 for WBC/

Table 1
Samples used in the study.

Samples	Cell type	Sex	Age, years	Source	Disease	Coriell ID	Platform used	Frozen before DNA extraction
WM1	WBC	M	26	HSC[a]	Healthy	N/a	Agilent, Illumina	N/A
LM1_LP	LP_LCL	M	26	HSC	Healthy	N/a	Agilent, Illumina	No
LM1_FR	LP_LCL	M	26	HSC	Healthy	N/a	Illumina	Yes
WM2	WBC	M	35	HSC	Healthy	N/a	Illumina	N/A
LM2_LP	LP_LCL	M	35	HSC	Healthy	N/a	Illumina	No
LM2_FR	LP_LCL	M	35	HSC	Healthy	N/a	Illumina	Yes
WF1	WBC	F	33	HSC	Healthy	N/a	Illumina	N/A
LF1_LP	LP_LCL	F	33	HSC	Healthy	N/a	Illumina	No
LF1_FR	LP_CLC	F	33	HSC	Healthy	N/a	Illumina	Yes
WF2	WBC	F	28	HSC	Healthy	N/a	Illumina	N/A
LF2_LP	LP_LCL	F	28	HSC	Healthy	N/a	Illumina	No
LF2_FR	LP_LCL	F	28	HSC	Healthy	N/a	Illumina	Yes
LM3_HP	HP_LCL	M	13	Coriell[b]	Healthy	GM06998	Illumina	Yes
LF3_HP	HP_LCL	F	12	Coriell	Healthy	GM12244	Agilent, Illumina	Yes
LM4_HP	HP_LCL	M	10	Coriell	Healthy	GM03798A	Illumina	Yes
WF4	WBC	F	0	HSC	Healthy	N/a	Agilent	N/A
LPID11_HP	HP_LCL	F	1	HSC	UPD11. BWS	N/a	Agilent, Illumina	Yes
LF5_LP	LP_LCL	F	7	HSC	Healthy	N/a	Pyro	No
WF5	WBC	F	7	HSC	Healthy	N/a	Pyro	N/A

For sample names the 1st letter indicates cell type (L for lymphoblastoid cell line, W for white blood cells), the 2nd letter indicates sex (M for male, F for female), and the third digit is individual number. If the number is the same, it means the sample originated from the same individual. The exception is LPID11 sample, which is a lymphoblastoid cell line having paternal isodisomy for chromosome 11. LP means low passage, HP means high passage, FR means the cell line was frozen before DNA extraction. N/A means not available.

[a] Hospital for Sick Children, TCAG, Biobanking facility (Toronto).

[b] NGIMS Human Genetics Cell Repository, Coriell Institute of Medical Research (Camden, NJ).

LP_LCLs and 0.940–0.954 for WBC/FR_LCLs. These results show that methylation profiles in LCLs are not identical to the original WBC samples, either due to the EBV transformation or to the different cell composition, as only B lymphocytes are transformed into LCLs. The correlation is very similar for WBCs/LP_LCLs and WBCs/FR_LCLs pairs, suggesting that one cycle of freezing does not affect genome-wide DNA methylation any more than short term culturing.

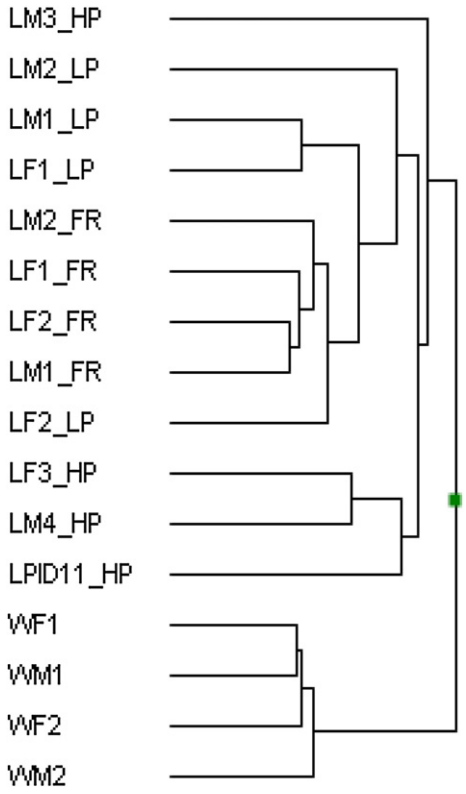

Fig. 1. Manhattan hierarchical clustering of Illumina methylation data from WBCs and LCLs. Dendogram shows the 2 main branches, WBCs and LCLs.

We hypothesized that if DNA methylation patterns changed at the same genomic locations in all LCLs, the correlation between LCLs would be very similar to the correlation among WBCs. However, we observed small but significant decreases of correlation coefficients between LP_LCL samples ($R = 0.960–0.982$, p-value 0.002, paired t-test) and FR_LCL ($R = 0.976–0.984$, p-value 0.013) compared to WBC samples ($R = 0.985–099$). Thus, these data indicate that EBV transformation and short-term culture can induce random DNA methylation changes. However one cycle of freezing does not induce increased methylation differences compared to short term culture only.

We further tested DNA methylation in the 4 HP_LCLs that were cultured for a higher number of cell divisions than low passage/frozen cell lines. We did not have matching DNA samples from WBCs for the HP_LCLs and were not able to track how many cell divisions these cell lines had undergone or how many times they were frozen prior to DNA extraction. Correlation coefficients for the pair-wise comparisons of DNA samples from HP_LCLs were in range of 0.905–0.955 (Table 2A). This is significantly lower than the correlations for WBCs (p-value = 0.0008) and LP_LCLs (p-value = 0.004, un-paired t-test). This reduction in correlation indicates that prolonged cell culturing and repeated freeze–thaw cycles induce progressive random changes in DNA methylation.

Agilent/MeDIP methylation correlations

MeDIP and Agilent CpG island microarray was performed on 5 samples WM1 and LM1_LP, WF4 (WBC DNA from a new-born female) and LF3_HP and LPID11_HP (Table 1). We performed pair-wise comparisons of all microarray data points in quantile normalized log2 ratios of MeDIP/input. Linear-regression correlation coefficients (R) are shown in Table 2B.

We observed the same trend in this experiment as for the Illumina interrogation. The methylation correlation among two WBC samples run on the Agilent microarray was higher or equal but not lower than correlations between LCL samples. The inter-individual correlation for the WM1_WF4 sample pair run on Agilent was slightly lower (0.972) than for WBC samples (WM1, WM2, WF1, WF2) run on the Illumina microarray (0.985–0.990). This could reflect specific methylation differences of one particular sample WF4, as this sample was collected from a newborn cord blood, whereas the other WBC samples were from adults.

Table 2
Correlation coefficients (R) of pair-wise comparisons of all data points (excluding sex chromosomes) between samples run on the Illumina array (A) and Agilent array (B).

A	LM1_LP	LM1_FR	WM2	LM2_LP	LM2_FR	WF1	LF1_LP	LF1_FR	WF2	LF2_LP	LF2_FR	LM3_HP	LF3_HP	LM4_HP	LPID11_HP
WM1	0.947	0.950	0.991	0.915	0.928	0.989	0.925	0.944	0.988	0.931	0.939	0.875	0.872	0.901	0.898
LM1_LP		0.983	0.949	0.962	0.973	0.933	0.982	0.978	0.948	0.969	0.975	0.928	0.943	0.966	0.950
LM1_FR			0.958	0.965	0.979	0.943	0.979	0.986	0.961	0.976	0.986	0.934	0.941	0.961	0.943
WM2				0.933	0.940	0.985	0.948	0.951	0.990	0.943	0.948	0.888	0.880	0.909	0.901
LM2_LP					0.983	0.910	0.965	0.961	0.930	0.960	0.962	0.920	0.927	0.949	0.929
LM2_FR						0.920	0.973	0.984	0.941	0.967	0.978	0.928	0.940	0.961	0.945
WF1							0.935	0.940	0.987	0.927	0.933	0.871	0.864	0.891	0.887
LF1_LP								0.987	0.947	0.971	0.977	0.931	0.944	0.965	0.949
LF1_FR									0.954	0.972	0.984	0.931	0.941	0.962	0.949
WF2										0.948	0.954	0.891	0.880	0.910	0.903
LF2_LP											0.987	0.931	0.932	0.956	0.934
LF2_FR												0.933	0.938	0.961	0.944
LM3_HP													0.909	0.927	0.905
LF3_HP														0.956	0.932
LM4_HP															0.947

B	WM1	LF3_HP	WF4	LPID11
LM1_LP	0.969	0.933	0.942	0.972
WM1		0.930	0.972	0.955
LF3_HP			0.905	0.920
WF4				0.934

WBC samples are highlighted in black, LP_ LCLs are highlighted in white, FR_LCLs are highlighted in light grey, and HP_ LCLs are highlighted in dark grey.

Detection of known methylation differences using Illumina and Agilent platforms

Known DNA methylation abnormalities in two imprinting centers were used to verify the capability of the Illumina and Agilent platforms to detect methylation changes. We used a DNA sample from a LCL with paternal isodisomy of chromosome 11 in 85–90% of the cells (LPID11_HP). This cell line from a patient with Beckwith–Wiedemann syndrome was known to carry methylation abnormalities at two imprinting centers on chromosome 11p15.5, gain of methylation at the paternal H19 DMR and loss of methylation at the maternal KvDMR [10]. We observed the expected loss of methylation at KvDMR on both the Illumina and Agilent arrays (Figs. 2A and C). The H19 DMR maps 2–4.4 kb upstream of the H19 transcription start site (TSS) (Imprinted Gene Catalogue Records: http://igc.otago.ac.nz). This particular region is not represented on either array, due to the presence of tandem repeats within the H19 DMR [13]. The H19 promoter is also known to be differentially methylated in human WBCs [14]. There are two CpG sites on the Illumina microarray, located between the promoter and the DMR, and as expected gain of methylation was observed at these sites (Fig. 2B) in the LPID11_HP sample compared to controls.

Interestingly, the high passage cell line from the healthy individual LM4_HP showed reduced methylation (β value = 0.18) at one of the CpG sites upstream of the H19 transcription start site (genomic location 1977136, build 36), whereas all other samples from WBCs of healthy individuals had β values ranging from 0.4 to 0.46, indicating that methylation at CpG sites within DMRs can be altered in high passage LCLs (Fig. 2B).

Methylation patterns at imprinted DMRs in LCLs

We analyzed known DMRs involved in the regulation of imprinted gene expression for changes in methylation patterns in LCLs from normal individuals, using both the Illumina and Agilent data. We checked 19 genomic regions that harbor known/potential DMRs [13,15–25]. The coordinates of the DMRs were taken from the Imprinted Gene Catalogue Records (http://igc.otago.ac.nz), when available, or were deduced from published data (Supplementary Table 1). The Illumina Infinium methylation array is reported to be accurate for β-value differences ≥ 0.2 (20%). Therefore this value was chosen as cut off for methylation differences. We set cut off levels of log2 ratio = 0.6 (1.5 fold change in methylation) in at least two contiguous probes for the Agilent data. This methylation difference between WBC and LCL of the same individual was required in order for a methylation change at a DMR to be accepted. For HP_LCL samples, for which we did not have an original WBC sample, the same cut off level was used, but the comparison was between each HP_LCL and all WBC samples run on this platform. Our assumption was that the methylation levels in the WBC samples of healthy individuals, from which HP_LCLs originated, are similar to those in normal WBC samples used in this study (WM1, WF1, WM2, WF2, WF4).

The majority of the DMRs tested retained their methylation patterns in LCLs. Methylation changes meeting the cut offs described above were found in 5 genes. These were DIRAS3, H19, MEG3, MKRN3, NDN and NNAT (Tables 3, Supplementary Table 2). The regions with altered methylation detected by microarrays are shown in Figs. 2B, 3, Supplementary Fig. 1 and Supplementary Table 2. There were more methylation changes occurring in HP_LCLs than in LP_LCLs and

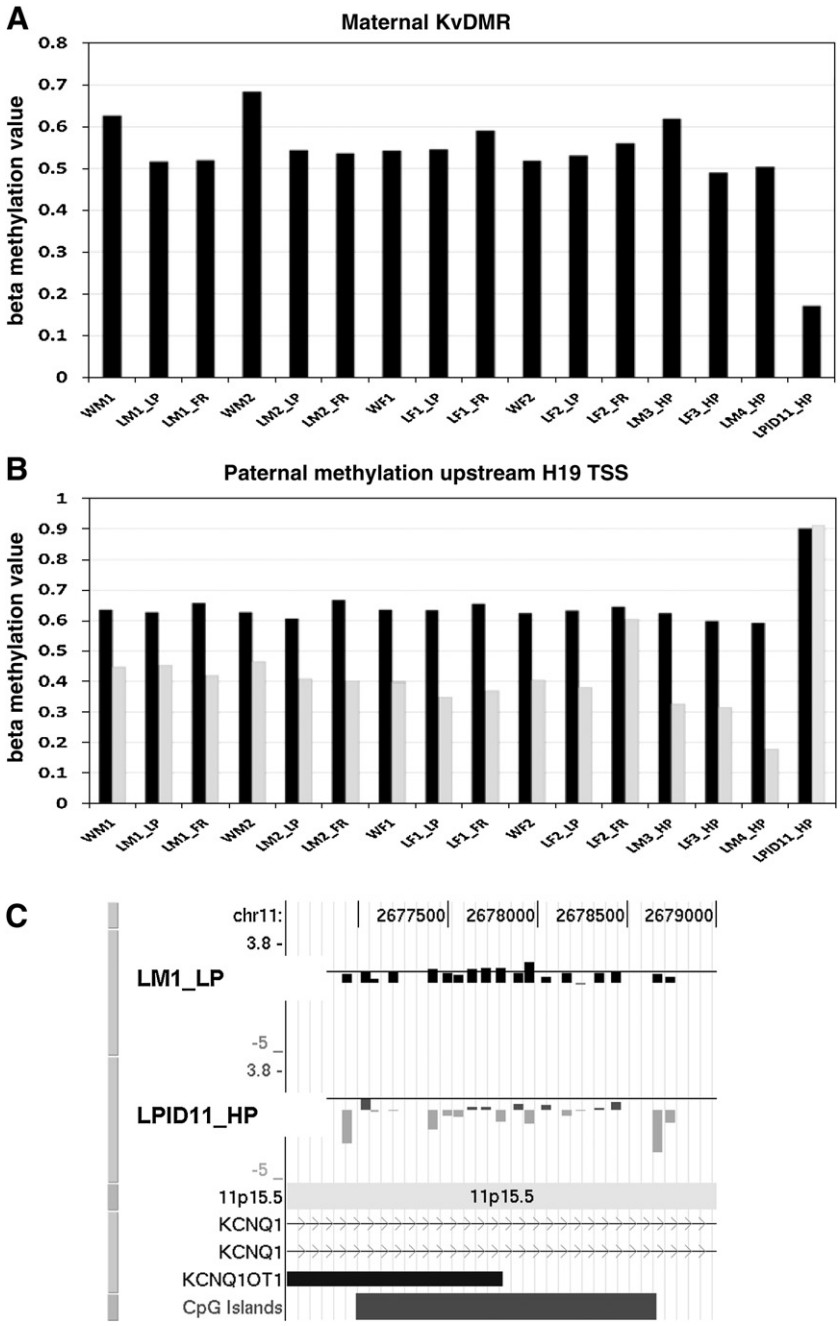

Fig. 2. Methylation array data for two DMRs at chromosome 11p15.5. (A) Illumina data for the maternal KvDMR CpG site (genomic coordinate chr11:2676805). (B) Illumina data for paternally methylated sites upstream of the H19 TSS (genomic coordinates chr11: 1976144 [black] and chr11: 1977136 [grey]. The *X* axis shows sample ID and the *Y* axis shows methylation values from 0 (not methylated to 1 (fully methylated). (C) Agilent CpG island microarray data integrated into UCSC Genome Browser for two samples: healthy control (LM1_LP) and patient with paternal UPD of chromosome 11 (LPID11_HP). The location of microarray probes within a CpG island is shown by black and grey bars. The probes are approximately 50 bp long. The height of each bar represents quantile normalized log2 ratios of methylated DNA/total DNA from hybridization signals for a single microarray probe. This corresponds to the level of methylation, i.e. higher methylation levels will correspond to higher bars. The LPID11_HP sample compared to control shows the expected reduction of methylation at KvDMR.

FR_LCLs. The methylation changes found in LCLs from healthy individuals were limited to two CpG sites and included both gains and losses. Gain of methylation was detected at 4 Illumina CpG sites and two Agilent probes for the NNAT promoter in the LPID11_HP sample from a BWS patient. This gain of methylation was detected by both arrays, because it occurred in several consecutive CpG sites (Supplementary Fig. 1).

To see if the DNA methylation changes detected in LCLs by the Illumina platform are limited to single CpG sites from the array, or expanded into surrounding CpGs, we designed pyrosequencing assays targeting the CpG site from the Illumina and several adjacent CpGs, for three genes: DIRAS3, H19, and NNAT. These assays were used to test the samples that were run on the array and additionally one pair of WBC and low passage LCL: WF5 and LF5_LP. DNA methylation values detected by Illumina and pyrosequencing are shown in Fig. 3. For H19 a CpG site with loss of methylation in LM4_HP from the Illumina array and 4 distal CpGs were tested by pyrosequencing. We found that DNA methylation changes in LM4_HP were limited to only one CpG site from Ilumina array, whereas adjacent CpGs retained normal methylation patterns (Fig. 2B).

For the DIRAS3 a CpG site (CpG7, Fig. 3D) with gain of methylation in LF3_HP and LPID11_HP and 6 distal CpG sites were tested by

Table 3
Methylation changes at known DMRs found in LCLs.

Gene	Position of DMR in the human genome	Source of coordinates	IC	Parent of origin methylation	Agilent met. change	Illumina met. change	Pyro met. change
DIRAS3 exon1	chr1:68288822–68289107	Luo et al.2001	Unkn	Putative maternal (E)	No	Gain in 1 CpG site in LM3_HP, LF3_HP, LPID11_HP	Gain in 2 CpGs in LF3_HP, LPID11_HP, 1 CpG in LF5_LP, loss in 1 CpG site in LF3_HP
H19 DMR	chr11:1977641–1980041, 2–4.4 kb upstream H19 TSS	http://igc.otago.ac.nz	Yes	Paternal	N/A	Loss in 1 CpG site in LM4_HP (outside of IC)	Loss in 1 CpG site in LM4_HP
MEG3	chr14:100359582–100361600	Rosa et al. 2005	No	Paternal	No	Loss in 2 CpG sites in LF3_HP, 1 CpG site in LM2_LP, LM4_HP, LPID11_HP	*N/A*
MKRN3	Promoter/exon1	Horsthemke and Wagstaff 2008	No	Maternal	N/A	Loss in 3 CpG sites in LF3_HP, 1 CpG site in LF1_FR, LM3_HP, LM4_HP	*N/A*
NDN	Promoter/exon1	Horsthemke and Wagstaff 2008	No	Maternal	No	Loss in 2 CpG sites in LM3_HP, 1 CpG site in LM2_LP, LM2_FR, LF1_LP, LF1_FR, and all HP_LCLs	*N/A*
NNAT	chr20:35580503–35584711	Evans et al. 2001	Unkn	Putative maternal (M,E)	Gain in 2 probes in LPID11	Loss in 1 CpG in LM3_HP, gain in 2 CpGs in LF2_LP, 1 CpG in LF2_FR and LF3_HP, and 4 CpGs LPID11_HP	Gain in 3 CpGs in LF2_LP, LF2_FR, LF3_HP, LPID11_HP LF5_LP, loss in 3 CpGs in LM3_HP

Positions of DMRS were either taken from http://igc.otago.ac.nz, or deduced from the cited papers. If parent of origin methylation is unknown in human, it was designated as putative, either based on mouse homology(M) and/or pattern of expression of downstream gene (E). N/A means information is not available due to the lack of the probes on the array. Unkn means unknown, regarding whether the DMR is an imprinting centre (IC).

pyrosequencing. Gain of methylation was observed in the CpG7 site and one adjacent CpG in both LF3_HP and LPID11_HP. Additionally, gain of methylation of CpG7 was found in LM5_LP, and loss of methylation was found in CpG1 in LF3_HP (Fig. 3D).

For NNAT the CpG site with loss of methylation in LM3_HP and gain of methylation in LF2_LP and LF2_FR and two distal CpG sites were validated by pyrosequencing. We observed losses and gains of methylation concordant with array data in all three sites and all three cell lines (Fig. 3F). Additionally we observed gain of methylation in all three sites in LF3_HP and LPID11_HP, which was also seen in the Illumina array, although slightly smaller than the set cut off of 0.2 methylation difference. We also found that there was gain of methylation for all three sites in LF5_LP compared to its WBC sample.

In summary we found that only a single CpG out of 5 CpGs was susceptible to methylation change for H19, whereas 2 out of 7 and 3 out of 3 CpGs were susceptible to DNA methylation change for DIRAS3 and NNAT respectively. Therefore we conclude the effect of cell culturing can be variable in terms of the length of the affected region.

Differential methylation analyses of Agilent CpG island microarrays

We identified the methylation changes between LM1_LP and WM1 samples first on Agilent microarrays, then checked the methylation patterns in the identified regions on Illumina arrays or by pyrosequencing. We tested additional WBC/LP_LCL pairs in order to verify whether these DNA methylation changes occur in all or just a subset of the LCLs, i.e. are they site-specific or random?

To set the parameters for the detection of methylation changes on the Agilent arrays we used the known methylation differences in samples LPID11_HP and LM1_LP. The LPID11 sample has paternal isodisomy of chromosome 11 in 85–90% of the cells and thus partial loss of methylation (17%) at maternal KvDMR (Fig. 2C). We adjusted the Partek GS analysis parameters in order to detect the difference in KvDMR methylation in the LPID11 sample (17%) compared to the methylation level in the LM1 sample (52%). These parameters were set to detect a 3-fold (52%/17%) and higher methylation change in 5 or more contiguous probes.

Using these parameters we searched for methylation differences in the LM1_LP sample compared to the WBC sample of the same individual, WM1. We detected 41 altered genomic regions, of which three were hypomethylated and 38 were hypermethylated in LM1_LP compared to WM1 (Table 4).

To check whether methylation changes found in LM1_LP are random or site-specific we further validated three regions using bisulphite pyrosequencing in 5 pairs of WBCs and LP_LCLs. As well we analyzed 6 regions using the Illumina platform in 4 pairs of WBCs and LP_LCLs.

UNC93B1, HTR1A and hsa-mir-339 (MIR339) were the genes selected for pyrosequencing. Primers for pyrosequencing were designed within the region of the altered DNA methylation, targeting 3 to 5 CpG sites per assay. For each sample, methylation values were averaged for all CpG sites within the assay. Microarray data, together with pyrosequencing results, are shown in Fig. 4A and Supplementary Fig. 2A. As Agilent data do not provide the percentage of methylation, we could not make a direct comparison between the pyrosequencing and Agilent data. However, we observed the same direction of change in methylation on the validation by pyrosequencing. For the UNC93B1 and hsa-mir-339 regions, we observed loss of methylation in all 5 LCLs tested, and for HTR1A we observed gain of methylation in 4 out of 5 LCLs tested.

Probes within 6 genomic regions (FAM84A, HOXA9, DLX5, WT1, HSPA2, IRX5) detected as different in LP1_LP versus WM1 by Agilent arrays were also present on the Illumina array. Therefore Agilent and Ilumina data could be cross-validated. The coordinates of the Illumina CpG sites are shown in Table 4. All of these regions were found to be hypermethylated in LM1_LP by Agilent. We observed the same direction of methylation change on Illumina in the WM1/LM1_LP sample pair (Fig. 4B, Supplementary Fig. 2B). Five of the six gains of methylations found in LM1 did not occur in all 4 LCLs tested by Illumina (Fig. 4B, Supplementary Fig. 2, right panel). Thus, these gains of methylation occur in only a subset of LCLs, i.e. these changes in methylation are not site-specific. Only gain of methylation within the WT1 gene was present in all 4 LCLs tested.

Thus the majority of gains of methylation (6 of 7) detected in LCLs represent changes that occur at random genomic locations. To determine the potential shared functionality of the genes acquiring methylation in LM1_LP, we used GO stat ontology analysis to identify the ontology categories which might be significantly overrepresented (p-value <0.01). The overrepresented genes fall into 67 different GO categories, showing that genes with many variable functions demonstrate gain of methylation in LCLs. We also did not observe any pattern in terms of the location of the DNA methylation changes relative to the transcription start of the gene (upstream, downstream, exonic, intronic) (Table 4).

In contrast to gain of methylation, loss of methylation in LM1_LP was detected for only three genes. The methylation status of two of them was confirmed by pyrosequencing in 5 WBCs/LCLs pairs, indicating that loss of DNA methylation in LCLs could be more site-specific than gain of methylation.

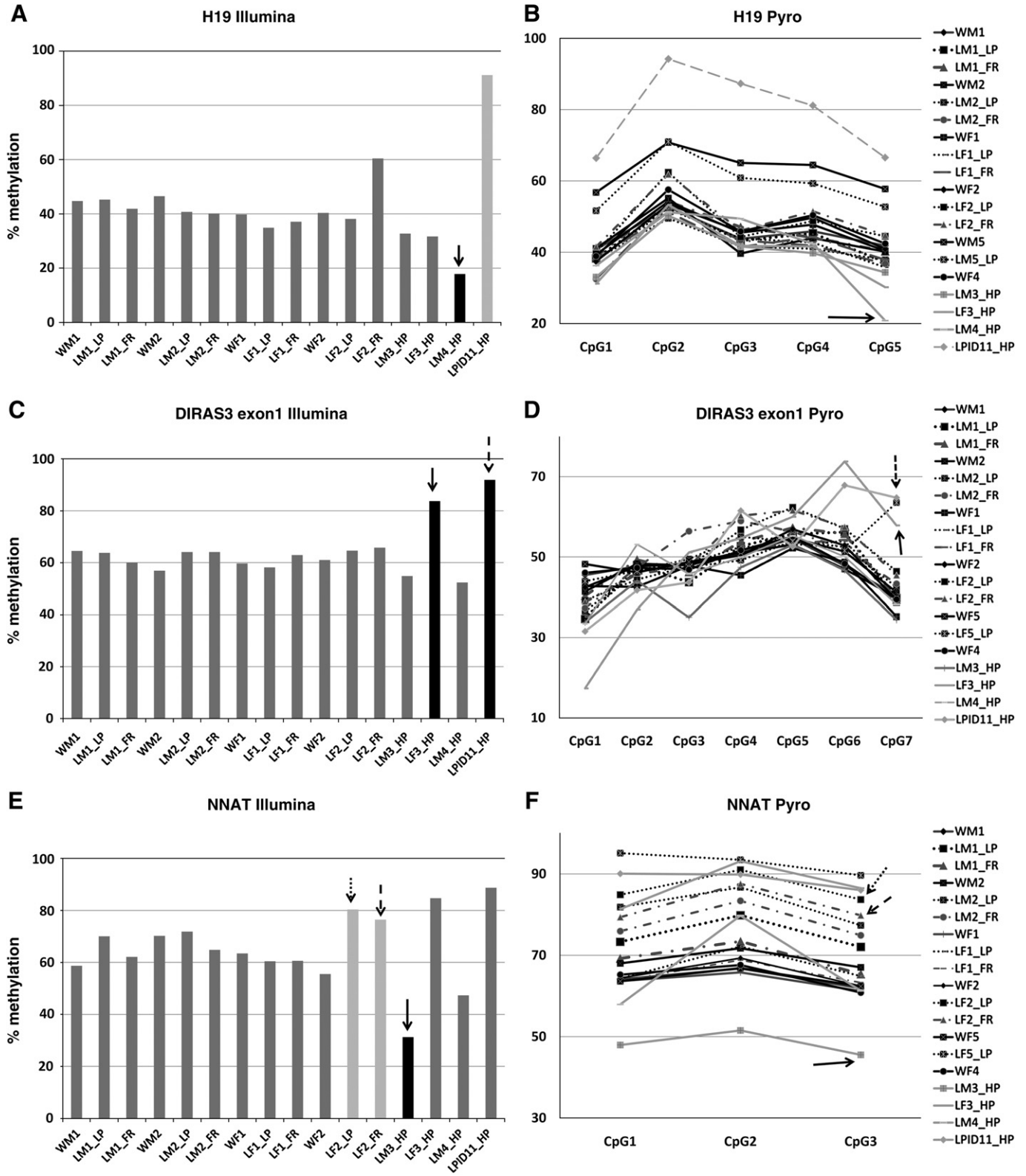

Fig. 3. DNA methylation profiles of single CpG site selected from the Illumina array analysis and corresponding pyrosequencing assay. For Illumina, β-values were transformed into % of methylation. The Illumina CpG sites within pyrosequencing assays are shown by arrows. The patterns of arrows are consistent between left panel (Illumina) and right panel (Pyrosequencing) to indicate the cell lines where a methylation change was found in Illumina array. The CpG sites within the pyrosequencing assay are ordered according to their position in the genome. Note that differences in % absolute values between Illumina and pyrosequencing could be due to amplification biases (towards amplification of unmethylated allele in pyrosequencing). (A) Illumina % DNA methylation at CpG site chr11:1977136 located upstream of H19 promoter TSS. (B) % DNA methylation determined by pyrosequencing at 5 consecutive CpG sites upstream of H19 TSS. (C) Illumina % DNA methylation at CpG site chr1:68288376 located in exon1 of DIRAS3. (D) % DNA methylation determined by pyrosequencing at 7 consecutive CpG sites in DIRAS3 exon1. (E) Illumina % DNA methylation at CpG site chr20:35582535 located in the NNAT gene. (F) % DNA methylation determined by pyrosequencing at 3 consecutive CpG sites in the NNAT gene.

Table 4
Genomic regions with differences in DNA methylation between WM1 and LM1_LP samples.

CHR	Start	End	Length, bp	State	# probes	Overlapping genes (±10 kb)	Position relative to the gene	Illumina
7	1029308	1029564	257	−4.7	5	C7orf50, hsa-mir-339	intron2 C7orf50, 5′ of hsa-mir-339	No
11	67526925	67527274	350	−4.7	5	UNC93B1, ALDH3B1	exon3 UNC93B1, 5′ of ALDH3B1	No
19	16297914	16299151	1238	−4.7	14	KLF2	exon1/intron1 KLF2	No
2	14690488	14691009	522	4.7	5	FAM84A	intron1 of FAM84A	14690908
2	119323820	119324191	372	4.7	6	EN1	5′ of EN1	No
2	176654328	176654697	370	4.7	5	EVX2	intron2 of EVX2	No
4	134291974	134292355	382	4.7	5	PCDH10	exon1 of PCDH10	No
4	174687959	174689038	1080	4.7	9	HAND2	5′ of HAND2	No
5	2803353	2803720	368	4.7	5	IRX2	intron1 of IRX2	No
5	3649976	3650264	289	4.7	5	IRX1	intron1 of IRX1	No
5	38592975	38593348	374	4.7	5	LIFR	intron1/5′ of isoforms of LIFR	No
5	63292612	63293068	457	4.7	6	HTR1A	exon1 of HTR1A	No
5	72712528	72712837	310	4.7	5	None	N/A	No
5	72776250	72776559	310	4.7	5	FOXD1	3′ of FOXD1	No
5	170668435	170668820	386	4.7	6	TLX3, RANBP17	5′ of TLX3, RANBP17	No
6	94184107	94184732	626	4.7	7	EPHA7	intron1 of EPHA7	No
6	101953405	101953889	485	4.7	7	GRIK2	5′ and exon1 of GRIK2	No
7	1236249	1237112	864	4.7	9	UNCX	5′ of UNCX	No
7	1245154	1245511	358	4.7	5	UNCX	3′ of UNCX	No
7	8449611	8449945	335	4.7	5	NXPH1	intron 2 NXPH1	No
7	27171824	27172255	432	4.7	5	HOXA7, HOXA9, HOXA10	5′ of HOXA7, HOXA9, 3′ of HOXA10	27172004
7	96489326	96489820	495	4.7	5	DLX5	exon/intron2 of DLX5	96489522
8	23619167	23619655	489	4.7	6	None	NA	No
10	131654094	131654647	554	4.7	7	EBF3	5′ of EBF3	No
11	20137724	20138119	396	4.7	6	DBX1	exon1/intron1 of DBX1	No
11	20138485	20138963	479	4.7	7	DBX1	5′ of DBX1	No
11	32411958	32412537	580	4.7	7	WT1	intron 1 of WT1	32412311
12	21985705	21986138	434	4.7	5	ABCC9	5′ of ABCC9	No
13	52320388	52320934	547	4.7	6	PCDH8	5′ and exon1 of PCH8	No
13	57105096	57105607	512	4.7	7	PCDH17	exon1 PCDH17	No
13	57105937	57106322	386	4.7	6	PCDH17	exon1 of PCDH17	No
13	111771263	111771606	344	4.7	5	SOX1	exon1 of SOX1	No
14	37747128	37747533	406	4.7	6	SSTR1	intron1 of SSTR1	No
14	56344670	56345360	691	4.7	8	OTX2	intron1 or 2 of isoforms of OTX2	No
14	64078497	64079033	537	4.7	6	ZBTB1, HSPA2, C14orf50	exon2 of HSPA2, 5′ of ZBTB1 and C14orf50	64079018
14	78815066	78815433	368	4.7	6	NRXN3	5′ or intron 5 of NRXN1 isoforms	No
15	72212252	72212678	427	4.7	7	ISLR2	exon1or 4 of isoforms of ISLR2	No
16	49745429	49746274	846	4.7	10	SALL1	5′ of SALL1	No
16	53522988	53523474	487	4.7	7	IRX5	5′ or intron1 of isoforms of IRX5	53522993
17	76930513	76931022	510	4.7	7	None	N/A	No
18	18002502	18002878	377	4.7	5	GATA6	5′of GATA6	No

Start and end length indicate start and end of the region showing changes in methylation. State: if positive shows gain of methylation in LM_LP and if negative loss of methylation in LM1_LP. Overlapping genes column shows the genes present within 10 kb of the detected region. The regions selected for pyrosequencing validation are shown in grey. The Illumina column shows the genomic coordinate of a CpG site. CHR is chromosome.

Discussion

In this paper, we compared genome-wide methylation profiles of WBC samples, low-passage LCLs pre and post freezing from four individuals and four unrelated high-passage LCLs. Our results show that DNA methylation profiles of LCLs and WBCs of the same individual are not identical (correlation coefficients = 0.933–0.947). It is in agreement with data of Brennan et al. [26], who studied differences in methylation of 318 genes using Sequenom's MassARRAY in 6 pairs of WBCs and LCLs, and found that 8% of these genes showed different methylation in LCLs compared to WBCs. This difference is not surprising, because LCLs originate from B-lymphocytes; the methylation profile for that cell type could be different from the pool of all white blood cells. Additionally EBV-transformation and cell culturing can potentially induce methylation changes. For example, Wilson and Jones [27] found that global methylation levels tend to decrease in fibroblasts over time in culture, whereas immortalized cell lines maintain more stable levels of methylation.

We further looked at the correlation of methylation patterns in LCLs to determine whether alterations occur at specific or random genomic locations. The correlation coefficients were very high between WBC samples (0.985–0.99) indicating very small inter-individual variation, but dropped significantly for LCLs pair-wise comparisons, which were especially profound for the cell lines with a history of multiple passages (Table 2). These data indicate that methylation changes occurring in the LCL do not affect all cell lines equally, e.g. a given genomic location can demonstrate altered methylation in one cell line but not in another. Surprisingly, the effect on methylation of one cycle of freezing was no stronger than that of just short term culturing.

We compared the regions with consistent changes in methylation across several probes in one WBC/LCL pair using the Agilent array, followed by validation of several regions using the Illumina array and pyrosequencing. This analysis revealed that the majority of gains of methylation (6 out of 7) detected in LCLs represent changes that do not occur in all WBCs/LCLs pairs and that they affect genes with variable functions. However, loss of methylation was detected in only three genes, and was shown for two of those genes to affect all WBCs/LCLs pairs. We suggest that loss of methylation detected in LCLs compared to WBCs might reflect differences in methylation between B-lymphocytes and other sub-types of WBCs, whereas gain of methylation represents random effects of cell culturing. In support of this, all three hypomethylated genes detected in LCLs by the Agilent platform have known functions in subtypes of white blood cells: UNC93B1 in B-cells [28], hsa-mir-339 in the human leukemic HL-60 cell line, which originates from peripheral white blood cells [29], and KLF2 in T-cells [30]. In contrast hypermethylated genes have variable gene functions.

For differentially methylated regions associated with imprinted genes, we observed changes in LP_LCLs, FR_LCLs and HP_LCLs in three imprinted genes MKRN3, NDN and NNAT. DIRAS3 and MEG3 were affected in LP_LCLs and HP_LCLs, but not in FR_LCLs, indicating again that one cycle of freezing has less effect on DNA methylation than culturing. H19 was affected in only one HP_LCL line. Interestingly, none of the affected imprinted regions are located within known imprinting centres. MEG3, NDN and MKRN3 harbor secondary DMRs dependent on imprinting centers within bigger imprinting clusters [23,25]. For single imprinted genes DIRAS3 and NNAT differential methylation levels were observed in somatic tissues [17,20]; however, it is unknown if these regions are primary ICs for these genes. The CpG site upstream of the H19 transcription start site showed loss of methylation in the HP_LCL sample. It is located outside of the region defined as the H19 DMR or IC1 of the 11p15.5 imprinted cluster [13]. These data suggest that ICs might be more resistant to methylation change in cell lines than other genomic regions.

The observed changes in LCLs indicate that these cells may not be an optimal screening tool for identification of putative imprinted DMRs by DNA methylation analysis, as DNA methylation levels could shift from the expected 50% due to cell culturing. The expression of imprinted genes could also potentially shift from monoallelic to biallelic due to methylation changes in LCLs or vice versa. As approximately 10% of human genes exhibit random patterns of monoallelic expression within single clones of LCLs [31], monoallelic expression in LCL could lead to a false conclusion that a gene is monoallelically expressed due to imprinting, when in fact it would be biallelic in a sub-population of these cells.

Overall, DNA methylation analysis in WBCs and three types of LCLs, in imprinted DMRs and genome-wide, showed that two initially very similar methylation patterns can become more and more different as the cell line undergoes multiple cell divisions.

A variety of mechanisms could induce random methylation changes in LCLs. One possibility is a model proposed by Antequera et al. [32] that any gene with an associated CpG island has a significant probability of becoming inactivated by DNA methylation, and if this occurs for a gene essential for survival, the cell will die. Alternatively when the gene is not essential for survival in cell culture, the cell survives carrying the altered methylation pattern, and may even have a selective advantage [32]. This model explains gain of methylation rather than loss and is consistent with our data that demonstrate more frequent random gain of methylation in LCLs.

Another possible reason for non-specific changes in DNA methylation could be clonality of the LCLs. Plagnol et al. examined X inactivation skewing as a measure of clonality in 1174 female LCLs, and found near monoclonality in 20% of the cell lines [33]. Also Simon-Sanchez et al. [1] observed that the ratio of cells with non-germ line heterosomic deletions was reduced in LCLs compared to the original WBC samples, probably due to the putative selective advantage of cells carrying the full diploid genome. The same mechanism could be true for methylation. If the initial population of cells was mosaic for a methylation change in a specific gene, a clonal increase for this specific change could shift the percentage of methylation. If this mechanism is important, the number of non-specific changes should accumulate with increases in cell divisions, which is what we observed.

In summary, we have shown that lymphoblastoid cell lines are prone to DNA methylation changes at random genomic locations

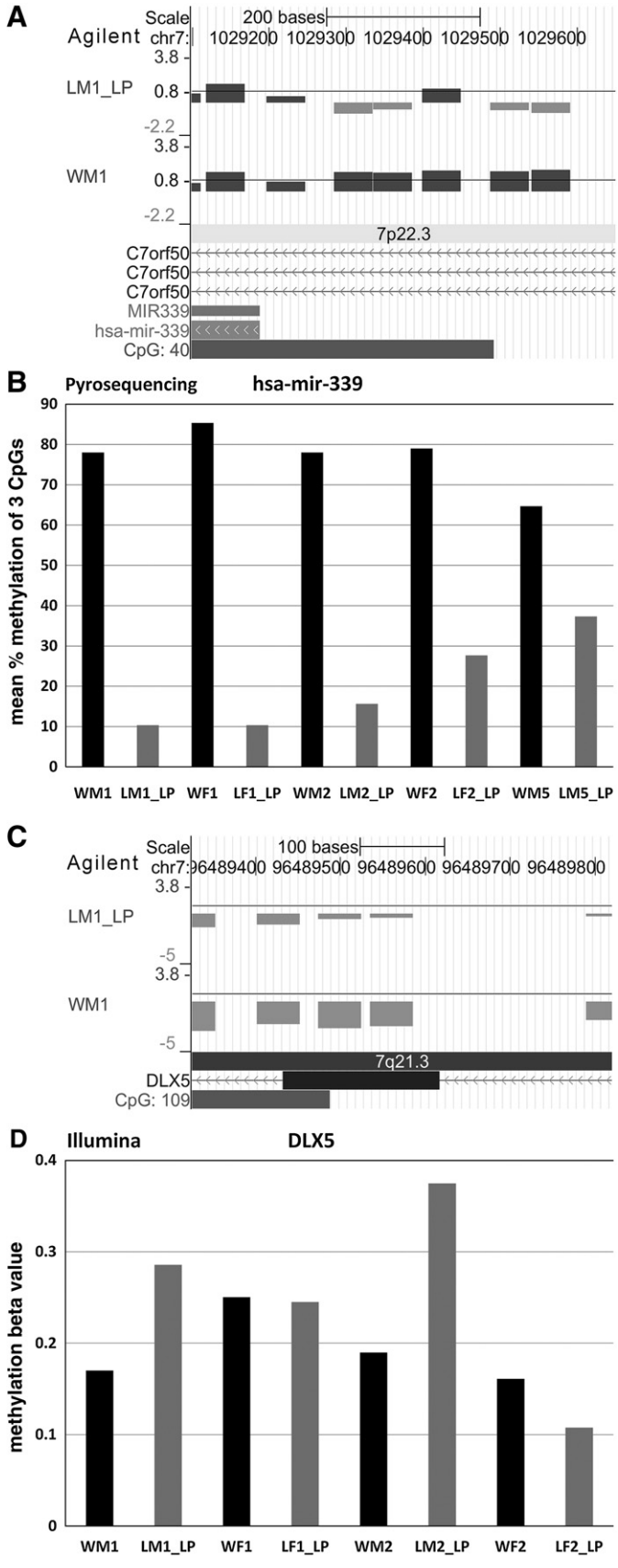

Fig. 4. (A) Agilent CpG island microarray data integrated into the UCSC Genome Browser for two samples: LM1_LP and WM1 within hsa-mir-339 promoter. The height of each bar represents quantile normalized log2 ratios of methylated/total DNA from hybridization signals for a single microarray probe. This corresponds to the level of methylation, i.e. higher methylation levels will correspond to higher bars. The line is drawn at log2 ratio = 0.8, which approximately equals 50% methylation. (B) Validation of array data within hsa-mir-339 promoter by pyrosequencing in 5 pairs of WBCs and LP_LCLs. (C) Agilent data for region of hypermethylation in the LM1_LP sample compared to the WM1 sample within DLX5 gene. (D) Illumina data for 1 CpG site within detected DLX5 gene for 4 pairs of WBCs and LP_LCLs.

especially after a high number of passages. Thus, special care should be taken when identifying disease-associated DNA methylation variants in LCLs. Confirmation of such alterations should be carried out in primary cell types.

Materials and methods

Research subjects

The study was approved by the Research Ethics Board of the Hospital for Sick Children (Toronto). Informed consents were obtained from all individuals.

Lymphoblastoid cell line establishment

Lymphoblastoid cell lines were established in the Biobanking facility, The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada according to standard procedures [34] or purchased from NGIMS Human Genetics Cell Repository, Coriell Institute of Medical Research (Camden, NJ) (http://ccr.coriell.org/).

DNA samples for microarray and pyrosequencing

Four matched sets of white blood cells (WBC) and low passage lymphoblastoid cell lines pre (LP_LCL) and post freezing (FR_LCL) from the same individuals were used in this study for LCL/WBC DNA methylation comparison by microarray. One additional pair of WBC/LP_LCL was used for pyrosequencing. Four unmatched high passage LCLs (HP_LCL) and one WBC sample were used for microarrays. A cell line with paternal isodisomy for chromosome 11 was used to validate microarray data and to optimize the Agilent microarray analysis parameters. For LP_LCLs cells were cultured after EBV transformation to 4×10^7 cells, 4 vials of $\sim8\times10^6$ cells were frozen and $\sim5\times10^6$ cells were cultured to 2×10^7 cells (which is equivalent to growing up to 16×10^7 cells if no freezing is done). For FR_LCLs, 8×10^6 cells were thawed and grown up to 2×10^7 cells. For HP_LCLs, it was not possible to track the previous history of growth/freezing.

DNA was extracted from 2×10^7 LCLs and 5 ml of whole blood using phenol-chloroform and ethanol precipitation. Sample information, including cell type, age, sex, source and application used are listed in Table 1.

MeDIP and Agilent CpG island microarray

Analysis of DNA methylation profiles was performed using methylated DNA immunoprecipitation (MeDIP) followed by Human 244K Agilent CpG island microarray (Agilent, Santa Clara, CA) which contains 237,220 probes, covering 27,800 CpG islands (97.5 % of UCSC annotated CpG islands). The probes are 45–60 bp in length and are spaced at 100 bp from the middle of each probe to the next.

MeDIP was performed as initially described by Weber et al. [11] with slight modifications. Briefly, 10 μg of 5mC antibody (Eurogentec, San Diego CA) was incubated with 50 μl of Dynabeads ® M-28 Sheep anti-mouse IgG (Invitrogen, Burlington, ON) for 5 h in IP buffer (10 mM Na phosphate pH7.0, 140 mM NaCl, 0.05% Triton X-100) at 4 °C. Genomic DNA was sonicated to the sizes 200–1000 bp using the Sonicator® Ultrasonic Liquid Processor, Model XL2000 (Misonix, Farmingdale, NY). 4 μg of DNA was added to the Antibody-beads complex and incubated overnight at 4 °C. The DNA-Antibody-Dynabeads complex was washed three times with IP buffer, and incubated in TE pH 8.0, 1% SDS solution with 5 μl of Protein K (10 mg/ml) for 2 h at 55 °C. DNA was further purified using the Qiaquick PCR purification kit (Qiagen, Mississauga, ON).

Immunoprecipitated DNA (21 μl) and 250 ng of sonicated genomic DNA of the corresponding sample of MeDIP DNA was labeled using BioPrime® Array CGH Genomic Labeling (Invitrogen) Cyanine 3-dCTP and Cyanine 5-dCTP (Perkin Elmer, Waltham, MA) respectively.

Purification of labeled products, array hybridization and scanning were performed at the Microarray facility, The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada according to the Agilent Microarray Analysis of Methylated DNA Immunoprecipitation Manual (Version 1.0, May 2008).

Data from scanned images (tiff) were extracted using Feature extraction software version 9.0 (Agilent), Human Agilent CpG version 20070207 grid file and CGH V4-91 protocol.

Analysis of Agilent CpG island microarray

Agilent microarray data were analyzed using the Partek Genomic Suite (PGS) software according to the algorithm developed by Sadikovic et al. [35]. The feature extraction files were imported into Partek GS as red (Cy5, MeDIP fraction) and green (Cy3, total genomic DNA fraction) processed signals. The data were normalized to baseline, by dividing red signal (MeDIP fraction) by green signal (total genomic DNA fraction), log2 transformed and quantile normalized for all the arrays. Thus, data were represented as log2 (methylated DNA/total DNA) after inter-array normalization for each probe of the array. Higher log2 ratios correspond to higher methylation levels. An ANOVA tool was used to calculate the fold change in methylation for each probe between samples. The Hidden Markov Model (HMM) was applied to the fold change of each probe to identify regions of contiguous probes with similar fold change in methylation. The parameters of HMM were adjusted in order to identify known methylation difference in the LPID11_HP sample at the maternally differentially methylated region KvDMR (11p15.5) compared to the LM1_LP sample. The level of methylation was 17% in LPID11_HP and 52% in LM1_LP sample, according Illumina array data. The following cut-off were used after optimization (min. probes 5, detection states −4.7, 4.7, ignore state 0, max. probability 0.99, genomic decay 10,000, sigma 1) to compare WM1 and LM1_LP samples. These parameters were set to detect approximately 3-fold (51/17) and higher methylation changes in 5 or more contiguous probes.

The detected regions of differences in methylation were annotated using refFlat.txt.gz file (overlap 10, 000 upstream and downstream) downloaded from the UCSC genome browser, and containing the most updated annotation of RefSeq genes. The microarray data were visualized in the UCSC genome browser by creating wig files for each chromosome containing normalized log2 (methylated DNA/total DNA) ratios for each sample.

Illumina's infinium HumanMethylation27 BeadChip

The HumanMethylation27 BeadChip (Illumina, San Diego, CA) contains 27,578 individual CpG sites covering 14,000 genes. Methylation status of the interrogated CpG site is measured as the ratio of signal from a methylated probe (C not converted to T) relative to the sum of both methylated (C not converted to T) and unmethylated probes (C bisulfite-converted to T). 1 μg of genomic DNA was bisulfite converted with Imprint® DNA Modification Kit according the manufacturer's protocol (Sigma, Oakville, ON). Whole genome amplification, labeling, hybridization and scanning were performed at the Genetic Analysis facility, The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada.

Analysis of Illumina methylation data

The methylation ratios (β value = C/(T+C)) were extracted using the Methylation Module in the Illumina Bead Studio after average normalization. This β value ranges continuously from 0 (unmethylated) to 1 (fully methylated). The correlations among samples and hierarchical clustering were performed in Bead Studio (Illumina). The number of detected probes (p-value<0.05) was very high for all arrays and varied between 99.3% and 99.9%.

Genome annotation

All genome coordinates used in this paper are according to Build 36 of the human genome annotation.

Bisulfite pyrosequencing

DNA methylation analysis by pyrosequencing were performed as described by Tost and Gut [36]. Pyrosequencing assays containing two PCR primers and one sequencing primer were designed to target 3–7 CpG sites using PSQ Design Software (Biotage, Uppsala, Sweden). One of the PCR primers had a universal tag which annealed to the universal biotinylated primer. Genomic DNA (1 μg) was bisulfite converted with Imprint® DNA Modification Kit (Sigma) according to the manufacturer's protocol and eluted in 20 μl of water. Bisulfite converted DNA (1 μl) was amplified using Hot-Start Taq-polymerase (Qiagen). Amplicons were analyzed on the Luc 96 pyrosequencer as specified by the manufacturer (Biotage), and % methylation was quantified as a ratio of C (methylated C, not converted to U) to T(not-methylated C converted to U) using Pyro Q-CpG Software (Biotage). The primer sequences are shown in Supplementary Table 3.

Acknowledgments

We thank the biobanking, microarray, and genetic facilities of The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada. We would like to acknowledge Rageen Rajendram for creating wig-files for the UCSC genome browser.

The work was supported by Genome Canada/Ontario Genomics Institute, the Canadian Institutes of Health Research (CIHR), the McLaughlin Centre for Molecular Medicine and the Canadian Institute of Advanced Research. Funding for DG was provided by the Autism Training Research Program (McGill University, Montreal). SWS holds the GlaxoSmithKline-CIHR Pathfinder Chair in Genetics and Genomics at the University of Toronto and Hospital for Sick Children.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2009.12.001.

References

[1] J. Simon-Sanchez, S. Scholz, H.C. Fung, M. Matarin, D. Hernandez, J.R. Gibbs, A. Britton, F.W. de Vrieze, E. Peckham, K. Gwinn-Hardy, A. Crawley, J.C. Keen, J. Nash, D. Borgaonkar, J. Hardy, A. Singleton, Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals, Hum. Mol. Genet. 16 (2007) 1–14.
[2] R.A. Philibert, R. Crowe, G.Y. Ryu, J.G. Yoon, D. Secrest, H. Sandhu, A. Madan, Transcriptional profiling of lymphoblast lines from subjects with panic disorder, Am. J. Med. Genet. B Neuropsychiatr. Genet. 144B (2007) 674–682.
[3] Y. Nishimura, C.L. Martin, A. Vazquez-Lopez, S.J. Spence, A.I. Alvarez-Retuerto, M. Sigman, C. Steindler, S. Pellegrini, N.C. Schanen, S.T. Warren, D.H. Geschwind, Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of autism and reveals shared pathways, Hum. Mol. Genet. 16 (2007) 1682–1698.
[4] J.F. Costello, C. Plass, Methylation matters, J. Med. Genet. 38 (2001) 285–303.
[5] V.K. Rakyan, T.A. Down, N.P. Thorne, P. Flicek, E. Kulesha, S. Graf, E.M. Tomazou, L. Backdahl, N. Johnson, M. Herberth, K.L. Howe, D.K. Jackson, M.M. Miretti, H. Fiegler, J.C. Marioni, E. Birney, T.J. Hubbard, N.P. Carter, S. Tavare, S. Beck, An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs), Genome Res. 18 (2008) 1518–1529.
[6] M. Weber, I. Hellmann, M.B. Stadler, L. Ramos, S. Paabo, M. Rebhan, D. Schubeler, Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome, Nat. Genet. 39 (2007) 457–466.
[7] C.A. Iacobuzio-Donahue, Epigenetic changes in cancer, Annu. Rev. Pathol. (2008).
[8] B. Horsthemke, K. Buiting, Genomic imprinting and imprinting defects in humans, Adv. Genet. 61 (2008) 225–246.
[9] I.K. Temple, Imprinting in human disease with special reference to transient neonatal diabetes and Beckwith–Wiedemann syndrome, Endocr. Dev. 12 (2007) 113–123.
[10] C.A. Edwards, A.C. Ferguson-Smith, Mechanisms regulating imprinted genes in clusters, Curr. Opin. Cell Biol. 19 (2007) 281–289.
[11] M. Weber, J.J. Davies, D. Wittig, E.J. Oakeley, M. Haase, W.L. Lam, D. Schubeler, Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells, Nat. Genet. 37 (2005) 853–862.
[12] I. Keshet, Y. Schlesinger, S. Farkash, E. Rand, M. Hecht, E. Segal, E. Pikarski, R.A. Young, A. Niveleau, H. Cedar, I. Simon, Evidence for an instructive mechanism of *de novo* methylation in cancer cells, Nat. Genet. 38 (2006) 149–153.
[13] M.A. Frevel, S.J. Sowerby, G.B. Petersen, A.E. Reeve, Methylation sequencing analysis refines the region of H19 epimutation in Wilms tumor, J. Biol. Chem. 274 (1999) 29331–29340.
[14] L. Guo, S. Choufani, J. Ferreira, A. Smith, D. Chitayat, C. Shuman, R. Uxa, S. Keating, J. Kingdom, R. Weksberg, Altered gene expression and methylation of the human chromosome 11 imprinted region in small for gestational age (SGA) placentae, Dev. Biol. 320 (2008) 79–91.
[15] T. Arima, R.A. Drewell, K.L. Arney, J. Inoue, Y. Makita, A. Hata, M. Oshimura, N. Wake, M.A. Surani, A conserved imprinting control region at the HYMAI/ZAC domain is implicated in transient neonatal diabetes mellitus, Hum. Mol. Genet. 10 (2001) 1475–1483.
[16] L. Beatty, R. Weksberg, P.D. Sadowski, Detailed analysis of the methylation patterns of the KvDMR1 imprinting control region of human chromosome 11, Genomics 87 (2006) 46–56.
[17] H.K. Evans, A.A. Wylie, S.K. Murphy, R.L. Jirtle, The neuronatin gene resides in a "micro-imprinted" domain on human chromosome 20q11.2, Genomics 77 (2001) 99–104.
[18] J. Li, A.J. Bench, G.S. Vassiliou, N. Fourouclas, A.C. Ferguson-Smith, A.R. Green, Imprinting of the human L3MBTL gene, a polycomb family member located in a region of chromosome 20 deleted in human myeloid malignancies, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 7341–7346.
[19] D. Lucifero, C. Mertineit, H.J. Clarke, T.H. Bestor, J.M. Trasler, Methylation dynamics of imprinted genes in mouse germ cells, Genomics 79 (2002) 530–538.
[20] R.Z. Luo, H. Peng, F. Xu, J. Bao, Y. Pang, R. Pershad, J.P. Issa, W.S. Liao, R.C. Bast Jr., Y. Yu, Genomic structure and promoter characterization of an imprinted tumor suppressor gene ARHI, Biochim. Biophys. Acta 1519 (2001) 216–222.
[21] S.K. Murphy, A.A. Wylie, R.L. Jirtle, Imprinting of PEG3, the human homologue of a mouse gene involved in nurturing behavior, Genomics 71 (2001) 110–117.
[22] R. Ono, H. Shiura, H. Aburatani, T. Kohda, T. Kaneko-Ishino, F. Ishino, Identification of a large novel imprinted gene cluster on mouse proximal chromosome 6, Genome Res. 13 (2003) 1696–1705.
[23] A.L. Rosa, Y.Q. Wu, B. Kwabi-Addo, K.J. Coveler, V. Reid Sutton, L.G. Shaffer, Allele-specific methylation of a functional CTCF binding site upstream of MEG3 in the human imprinted domain of 14q32, Chromosome Res. 13 (2005) 809–818.
[24] M. Zeschnigk, B. Schmitz, B. Dittrich, K. Buiting, B. Horsthemke, W. Doerfler, Imprinted segments in the human genome: different DNA methylation patterns in the Prader–Willi/Angelman syndrome region as determined by the genomic sequencing method, Hum. Mol. Genet. 6 (1997) 387–395.
[25] B. Horsthemke, J. Wagstaff, Mechanisms of imprinting of the Prader–Willi/Angelman region, Am. J. Med. Genet. A 146A (2008) 2041–2052.
[26] E.P. Brennan, M. Ehrich, D.P. Brazil, J.K. Crean, M. Murphy, D.M. Sadlier, F. Martin, C. Godson, A.J. McKnight, D. van den Boom, A.P. Maxwell, D.A. Savage, Comparative analysis of DNA methylation profiles in peripheral blood leukocytes versus lymphoblastoid cell lines, Epigenetics 4 (2009) 159–164.
[27] V.L. Wilson, P.A. Jones, DNA methylation decreases in aging but not in immortal cells, Science 220 (1983) 1055–1057.
[28] I. Isnardi, Y.S. Ng, I. Srdanovic, R. Motaghedi, S. Rudchenko, H. von Bernuth, S.Y. Zhang, A. Puel, E. Jouanguy, C. Picard, B.Z. Garty, Y. Camcioglu, R. Doffinger, D. Kumararatne, G. Davies, J.I. Gallin, S. Haraguchi, N.K. Day, J.L. Casanova, E. Meffre, IRAK-4- and MyD88-dependent pathways are essential for the removal of developing autoreactive B cells in humans, Immunity 29 (2008) 746–757.
[29] S. Pizzimenti, M. Ferracin, S. Sabbioni, C. Toaldo, P. Pettazzoni, M.U. Dianzani, M. Negrini, G. Barrera, MicroRNA expression changes during human leukemic HL-60 cell differentiation induced by 4-hydroxynonenal, a product of lipid peroxidation, Free Radic. Biol. Med. (2008).
[30] A.F. Buckley, C.T. Kuo, J.M. Leiden, Transcription factor LKLF is sufficient to program T cell quiescence via a c-Myc-dependent pathway, Nat. Immunol. 2 (2001) 698–704.
[31] A. Gimelbrant, J.N. Hutchinson, B.R. Thompson, A. Chess, Widespread monoallelic expression on human autosomes, Science 318 (2007) 1136–1140.
[32] F. Antequera, J. Boyes, A. Bird, High levels of *de novo* methylation and altered chromatin structure at CpG islands in cell lines, Cell 62 (1990) 503–514.
[33] V. Plagnol, E. Uz, C. Wallace, H. Stevens, D. Clayton, T. Ozcelik, J.A. Todd, Extreme clonality in lymphoblastoid cell lines with implications for allele specific expression analyses, PLoS ONE 3 (2008) e2966.
[34] H. Neitzel, A routine method for the establishment of permanent growing lymphoblastoid cell lines, Hum. Genet. 73 (1986) 320–326.
[35] B. Sadikovic, M. Yoshimoto, K. Al-Romaih, G. Maire, M. Zielenska, J.A. Squire, *In vitro* analysis of integrated global high-resolution DNA methylation profiling with genomic imbalance and gene expression in osteosarcoma, PLoS ONE 3 (2008) e2834.
[36] J. Tost, I.G. Gut, DNA methylation analysis by pyrosequencing, Nat. Protoc. 2 (2007) 2265–2275.